УДК 004.93

Ю.В. Крак, Ю.Г. Кривонос, О.В. Бармак, А.С. Тернов
Інститут кібернетики імені В.М. Глушкова НАН України, м. Київ, Україна

Інформаційна технологія невербального спілкування людей з вадами слуху

Пропонується концепція комплексної інформаційної технології невербального спілкування українською мовою жестів людей з вадами слуху, як між собою, так і з іншими людьми.

Вступ і постановка задачі

У світі кількість повністю глухих людей і людей з важкими формами втрати слуху складає близько 1,5 % від загальної чисельності населення – це десятки мільйонів людей, для яких необхідно створювати засоби рівноцінного спілкування у суспільстві. Пункт 7 Правила 5 Додатку до Резолюції ООН 48/96 «Стандартні правила забезпечення рівних можливостей для інвалідів» говорить: «Потрібно подбати про те, щоб мова жестів застосовувалась для навчання глухих дітей, в їх сім'ях та громадах. Потрібно також надавати послуги по сурдоперекладу для того, щоб сприяти спілкуванню глухих з іншими людьми» [1].

В Україні більше півмільйона дітей з вадами слуху, а кількість глухих людей, для яких потрібно розробляти сучасні засоби навчання і спілкування відповідно до світового науково-технічного розвитку – становить мільйони. Розвиток сучасної науки, комп'ютеризація суспільства, використання мультимедійних та інтернет-технологій створили достатні умови для розробки комп'ютерних систем комунікації цих людей у формах і образах, близьких і зрозумілих для них і для оточуючого середовища.

Основною формою спілкування глухих є мова жестів. Мова жестів має національні особливості (наприклад, англійська, французька, українська та інші мови), причому люди з вадами слуху використовують у спілкуванні дві жестові мови, які мають різну граматику й різний набір жестів:
– розмовну мову жестів, яка використовується у повсякденному спілкуванні і має власну граматику, досить відмінну від природної розмовної мови;
– калькулюючу мову жестів, яка використовується в офіційній і діловій обстановці та містить у собі як знаки розмовної мови жестів, так і знаки дактилогічної абетки, яка служить для відтворення слів по буквах, причому калькулююча мова жестів не має власної граматики, вона підкоряється граматичним правилам національної мови.

Важливою проблемою спілкування глухих з іншими людьми є вміння розпізнавати розмовну мову по губах, оскільки звичайні люди загалом не знають і не вивчають мову жестів. З цього погляду задачі візуального синтезу мови та розпізнавання по губах є альтернативою мовного спілкування для людей з вадами слуху. Крім цього, розвиток напрямку автоматичного читання по губах допоможе

покращити показники існуючих систем розпізнавання мови завдяки отриманню додаткового незалежного каналу інформації, а її синтезоване озвучення дозволить включити в комунікаційний процес людей з вадами зору.

Огляд літератури з даної тематики показав, що на теперішній час подібні системи розроблені для англійської, французької мови [2], [3], початі дослідження для російської мови [4]. Подібні розробки для української мови відсутні.

Авторами обґрунтовується необхідність розробки комплексної інформаційної технології невербального спілкування українською мовою жестів людей з вадами слуху, як між собою, так і з іншими людьми. Виходячи з цього сформульована наступна постановка задачі.

Необхідно розробити концепцію комп’ютерної інформаційної технології невербального спілкування людей з вадами слуху (що не чують та слабо чують) з можливістю реалізації у її рамках наступної функціональності:
– синтез тексту українською мовою на природну голосову мову;
– відображення на обличчі тривимірної комп’ютерної моделі людини мовленнєвого процесу з урахуванням емоційних складових;
– синтез рухів мови жестів глухих та дактильної абетки тривимірної моделі людини;
– аналіз речень українською мовою для синтезу їх аналогів на мові жестів;
– аналіз міміки губ мовленнєвого процесу.

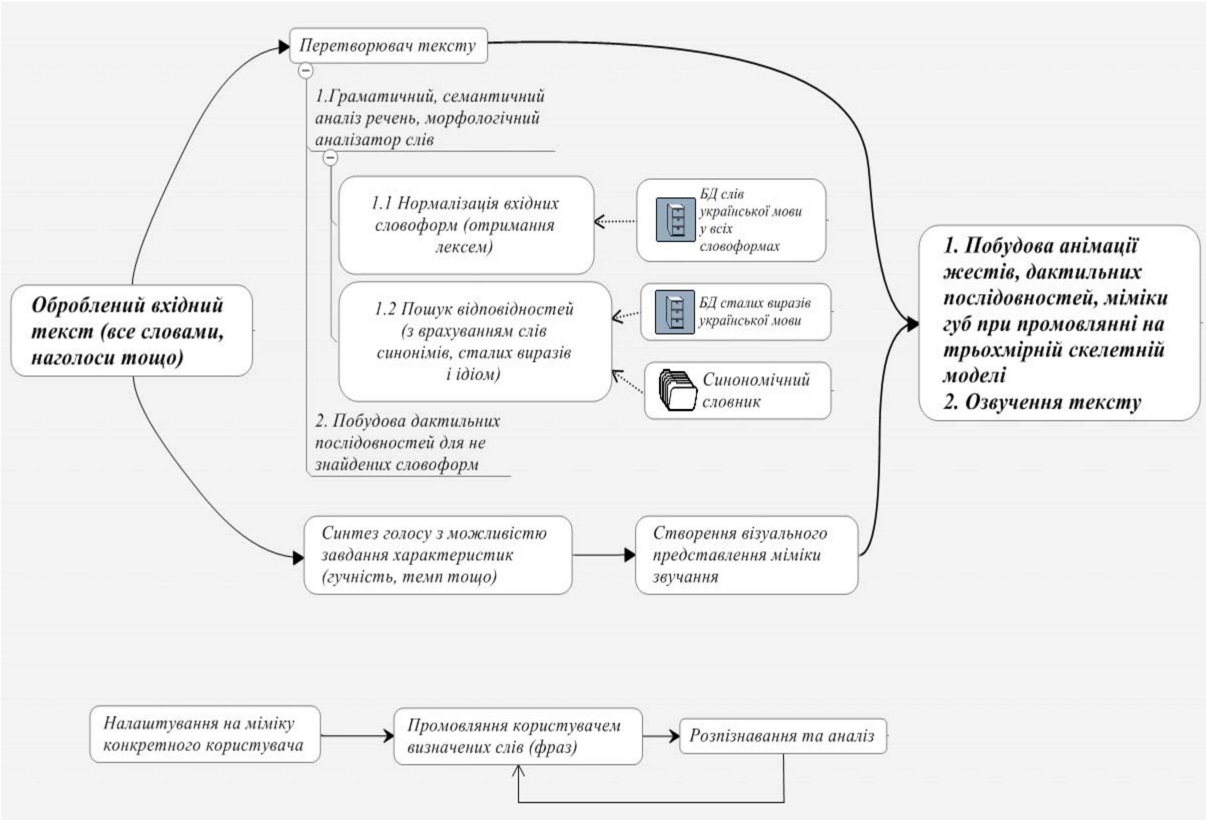
Інформаційна технологія невербального спілкування людей з вадами слуху

Комплексна інформаційна технологія включає реалізацію наступних можливостей:
– модуль перекладу звичайного тексту на мову жестів глухих, який міститиме анімацію процесу промовляння розмовної і калькулюючої мови жестів з використанням віртуальних моделей людей;
– анімацію міміки обличчя (з урахуванням емоційних складових) при промовлянні;
– озвучення (синтез) звичайного тексту у його звуковий аналог (з використанням різних голосів);
– модуль розпізнавання за зміною міміки губ тексту, який промовляється.

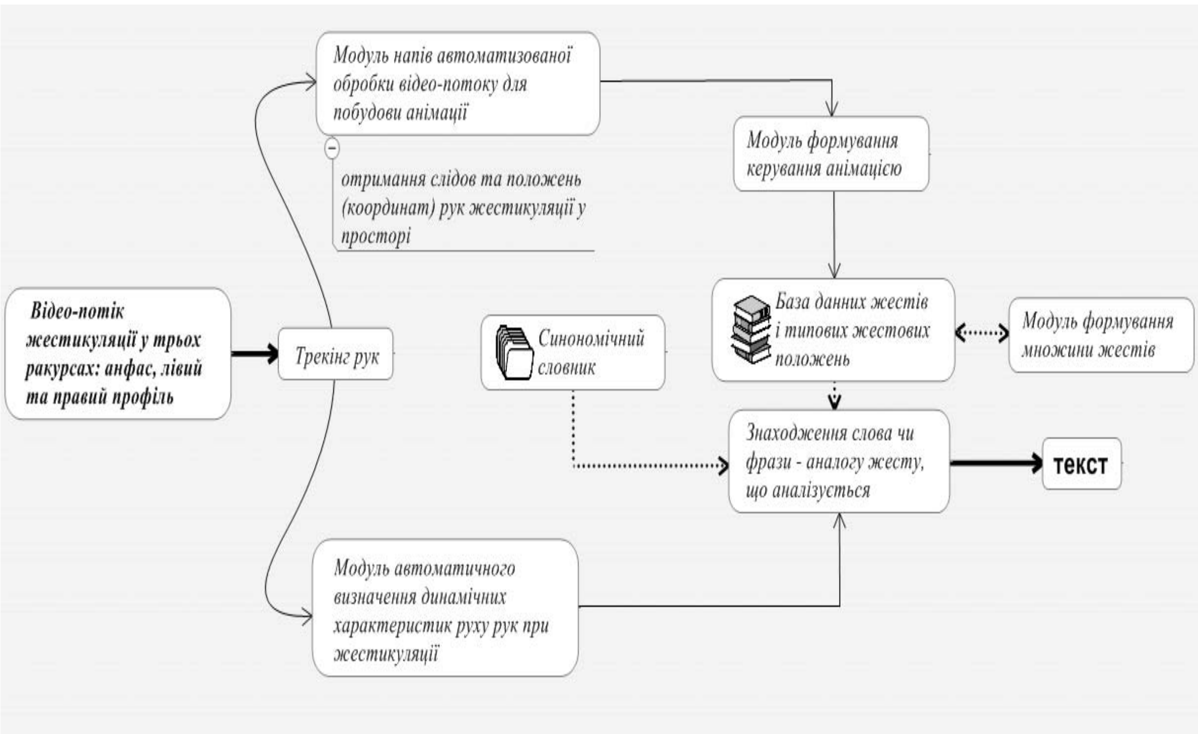
Концепція експериментальної технології віртуального спілкування людей з вадами слуху зображена на схемі рис. 1 а), b).

У перетворювачі тексту основне навантаження несе граматичний, семантичний аналіз речення та морфологічний аналіз слова. При перекладі речення мови жестів у речення українською мовою мають місце наступні проблеми:
– при нормалізації вхідної словоформи (морфологічний аналізатор) можливі ситуації, коли вхідна словоформа має декілька варіантів слів, що належать до різних частин мови. Для вирішення таких колізій необхідно попередньо провести граматичний аналіз речення, що суттєво зменшить такі колізії;
– з іншої сторони, можливі ситуації, коли одне і те ж слово української мови має багато відтінків та переходів в значеннях і вживанні. Тобто для точного визначення жесту для конкретного слова потрібно визначити семантичну структуру вхідного речення.

Схема алгоритму такої функції перетворювача зображена на рис. 2.



a)



b)

Рисунок 1 – Схема концепції експериментальної технології віртуального спілкування людей з вадами слуху та зору

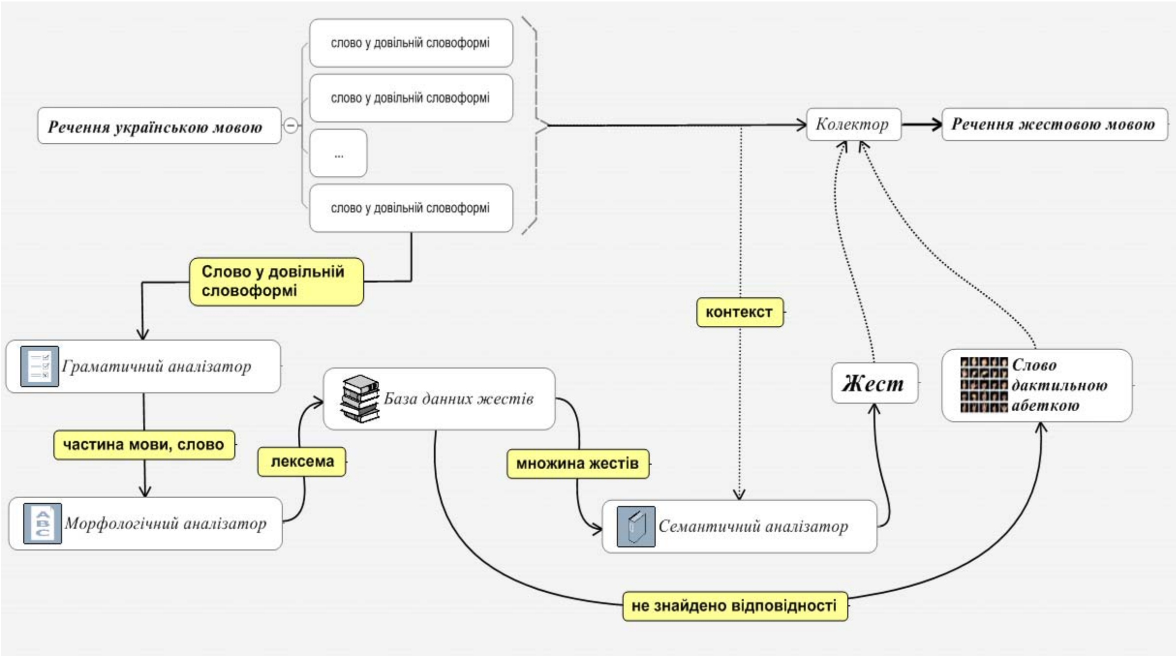


Рисунок 2 – Перетворювач тексту

Граматичний аналізатор визначає для кожного слова, до якої частини речення воно відноситься, і також частину мови, до якої дане слово належить.

Морфологічний аналізатор нормалізує вхідні словоформи (отримання лексеми).

Семантичний аналізатор, враховуючи значення слова та його вживання у даному реченні, визначає жест.

Реалізація концепції

В рамках реалізації запропонованої концепції комп'ютерної інформаційної технології невербального спілкування людей з вадами слуху проведені дослідження та розроблене тестове програмне забезпечення.

Для синтезу тривимірної анімації мови жестів побудовані геометричні класи векторів-образів жестів. Побудова цих класів базувалась на використанні технології motion capture [5]. Motion capture – це технологія подання рухів, яка дозволяє перейти від характеристик руху в реальному світі до фіксації змін параметрів для математичної моделі. Для автоматичного отримання необхідних ключових координат жесту використовувалась технологія трекінгу [6].

Для опису жесту використовувався формат BVH, з наступним експортом отриманих даних на скелетну тривимірну модель (наприклад, у модулі Character Studio для 3D studio MAX або в Poser).

Запропонована реалізація технології motion capture для фіксації рухів жестової мови містить наступні етапи (рис. 3):

1) отримання відеопотоку жестикуляції у двох ракурсах: фронт, профіль (усі необхідні параметри зйомки відомі);

2) обробка відеопотоку з виділенням слідів і положень (координат) рук жестикуляції в просторі;

3) на основі отриманих координат жестикуляції – формування BVH-файлу для синтезу трьохвимірної анімації;

4) застосування BVH-файлу для створення анімації (в Character Studio для 3D studio MAX або в Poser).

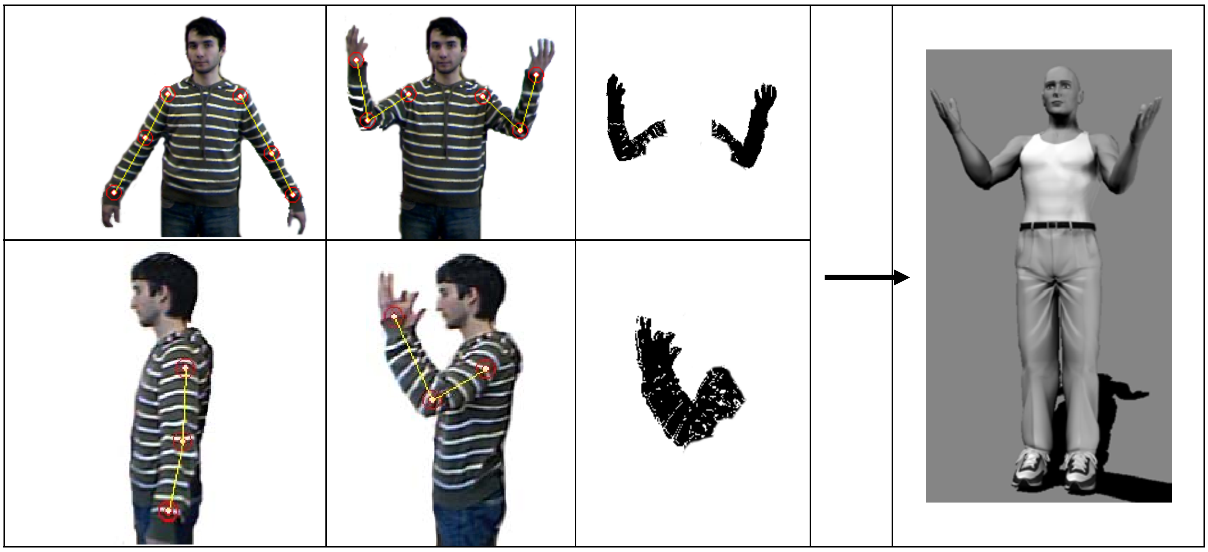


Рисунок 3 – Побудова анімації жесту

Для обробки вхідного тексту (розміщення наголосів, виділення інфінітивів, пошуку синонімів і типових фраз мови) розроблена інформаційна модель української мови. Модель подана у вигляді таблиць реляційної бази даних з набором збережених процедур, які реалізують необхідну для розглянутої технології функціональність (рис. 4).

Інформаційна модель української мови містить більше двох з половиною мільйонів слів (всі можливі словоформи), наголоси, синоніми, ідіоми, фрагменти слів для синтезу різними голосами, вектори-образи жестів.

Для реалізації функції візуалізації озвучення тексту розроблено синтезатор української мови. Синтезатор дозволяє створювати голосовий аналог довільного тексту різними голосами з можливістю керування характеристиками голосу (гучність, далі/ближче). Синтезатор також дає можливість побудувати візуальне подання процесу промовляння як за допомогою двомірних візем, так і на тривимірній моделі (рис. 5).

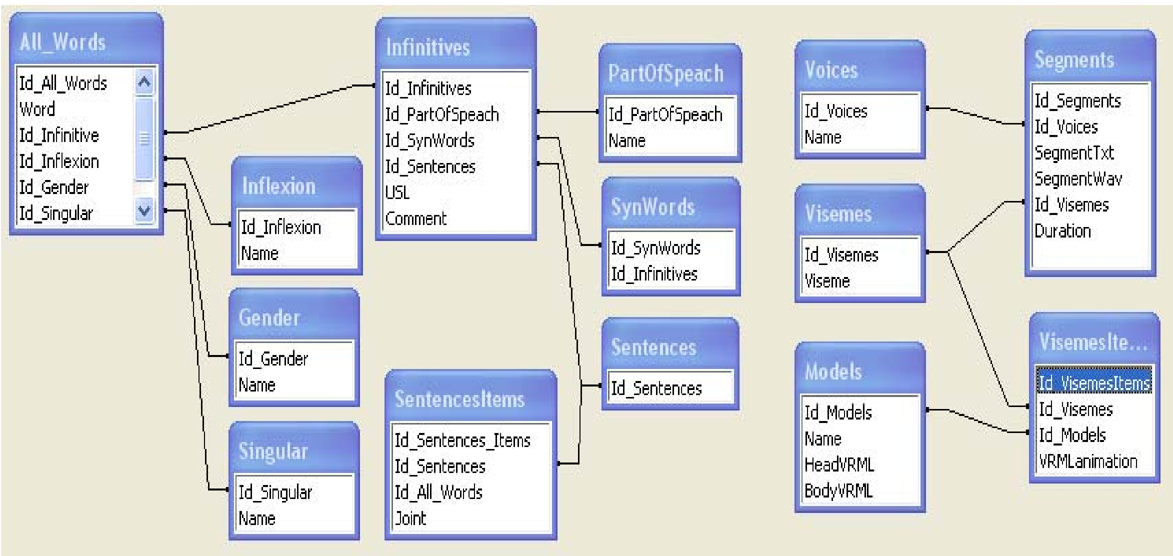
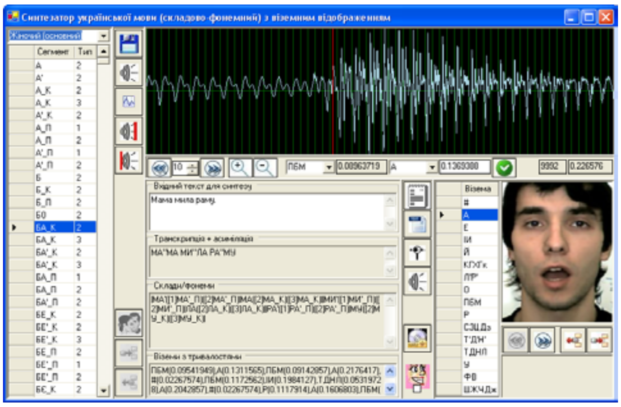
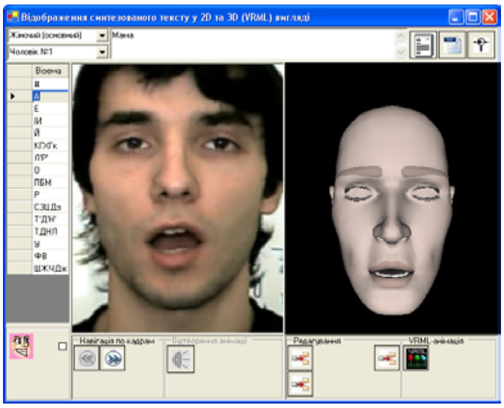


Рисунок 4 – Логічна схема бази даних – інформаційна модель української мови



a)



b)

Рисунок 5 – а) синтезатор української мови; b) синтез візем української мови

Для комплексного тестування запропонованої технології створене відповідне програмне забезпечення (рис. 6) для перекладу довільного тексту на калькулятивну мову жестів. Програмне забезпечення реалізує наступний алгоритм синтезу послідовності жестів:

1) вхідний текст синтезується у його голосовий аналог;

2) вхідний текст розбивається на слова;

3) для кожного слова у базі даних слів шукається його нормальна форма (інфінітив);

4) для кожної нормальної форми слова у базі даних жестів шукається відповідна послідовність рухів;

5) якщо для даної словоформи послідовність рухів не знайдена, то слово зображається дактильною абеткою;

6) тривимірна модель відтворює генеровані послідовності рухів та синтезований голосовий аналог.

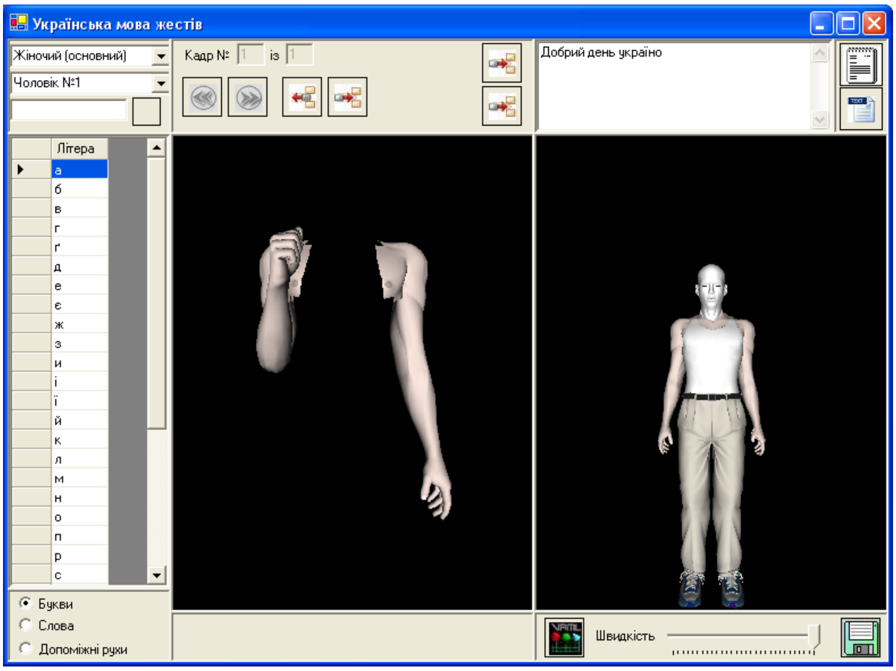


Рисунок 6 – Синтез мови жестів на об’ємній моделі

Висновки

В роботі запропонована концепція комплексної інформаційної технології невербального спілкування українською мовою жестів людей з вадами слуху, як між собою, так і з іншими людьми. Отримані результати в рамках цієї концепції показали її життєздатність та виявили деякі проблеми, які потребують подальших досліджень.

Подальші дослідження направлені на реалізацію всієї множини рухів української мови жестів. Окремого розгляду потребують питання, пов'язані з граматичним та семантичним аналізом речень.

Література

1. Стандартные правила обеспечения равных возможностей инвалидов // Офіційні звіти Генеральної Асамблеї, сорок восьмая сессия. – Дополнение N 49 (A/48/49). – С. 292-306. Документ 995_306, редакція від 20.12.1993. – Режим доступа: http://www.un.org/documents/ga/res/48/a48r096.htm
2. Neidle C., Sclaroff S., V. Athitsos. SignStream™: A Tool for Linguistic and Computer Vision Research on Visual-Gestural Language Data: Boston: Boston University, Massachusetts. – 2001. – 33:3. – P. 311-320.
3. IBM Research Demonstrates Innovative 'Speech to Sign Language' Translation System, Extreme Blue programme HURSLEY, UK – 12 Sep 2007, IBM Media Relations. – Режим доступа: http://www-03.ibm.com/press/us/en/pressrelease/22316.wss
4. Воскресенский А.Л., Хахалин Г.К. Мультимедийный толковый словарь русского жестового языка // Компьютерная лингвистика и интеллектуальные технологии: Труды международной конференции «Диалог 2007» (Бекасово, 30 мая – 3 июня 2007 г.) / Под ред. Л.Л. Иомдина, Н.И. Лауфер, А.С. Нариньяни, В.П. Селегея. – М.: Изд-во РГГУ, 2007. – 658 с.
5. Menache A. Understanding Motion Capture for Computer Animation and Video Games, Morgan Kaufmann. – 2000. – 238 p.
6. S. Avidan. Support vector tracking // Proc. IEEE Conf. on Computer Vision and Pattern Recognition, Kauai, Hawaii. – 2001. – Volume I. – P. 184-191.
7. Villani N., Doublestein J., Martin Z. The MathSigner: an interactive learning tool for American sign language Adamo-Information Visualisation, 2004. – IV. Proceedings. Eighth International Conference on Volume, Issue, 14-16 July. – 2004. – P. 713-716.

Ю.В. Крак, Ю.Г. Кривонос, А.В. Бармак, А.С. Тернов

Информационная технология невербального общения людей с недостатками слуха

Предложена концепция комплексной информационной технологии невербального общения на украинском языке жестов людей с недостатками слуха, как между собой, так и с другими людьми.

Yu.V. Krak, Yu.G. Kryvonos, A.V. Barmak, A.S. Ternov

Information technology for nonverbal communication of deaf people

Complex information technology for nonverbal communication of deaf people both between itself and with other people based on sign language is created.

Стаття надійшла до редакції 23.07.2008.